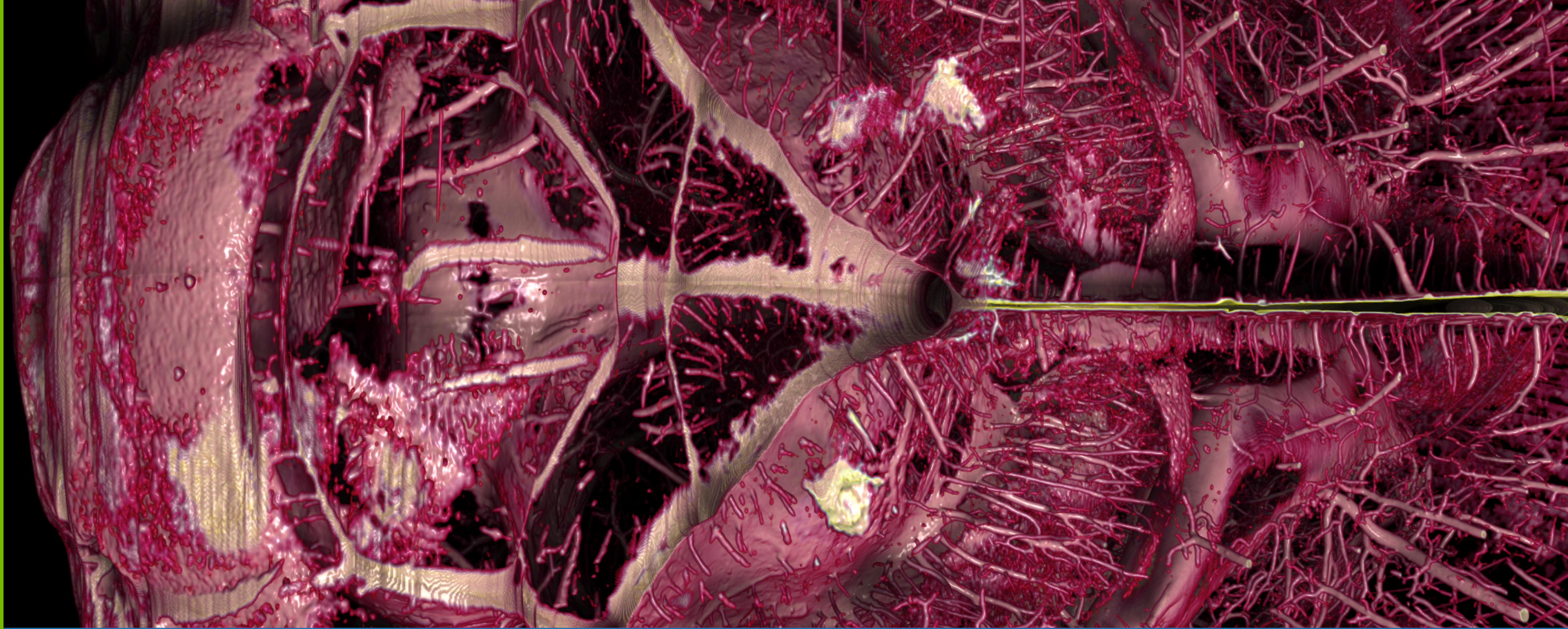


CONNECTOMICS EARLY SCIENCE PROJECT
(PI: NICOLA FERRIER, ARGONNE; BOBBY KASTHURI, UCHICAGO; SEBASTIEN SEUNG, PRINCETON)

THOMAS URAM, ARGONNE LEADERSHIP COMPUTING FACILITY
TURAM@ANL.GOV

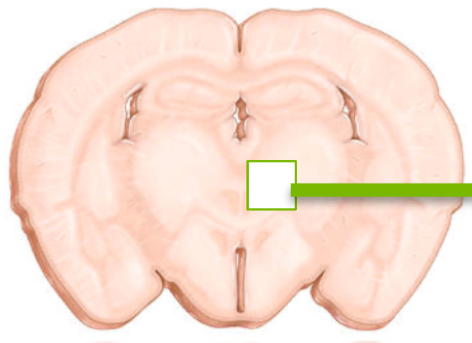
WHAT IS CONNECTOMICS?

- Connectomics attempts to reconstruct the neurons in a sample of brain tissue to obtain their shape and interconnectivity
- Why? To answer some fundamental and practical questions:
  - How does the brain achieve its functionality (sensory, motor, consciousness)
  - How do developmental changes (e.g. aging) affect the brain?
  - Is learning reflected in structure and connectivity?
  - How does plasticity help overcome brain injury?
  - Can we leverage an understanding of biological brains to improve neural networks?
  - Can we use biological structure to build faster and more power-efficient computers?

Data from Gregg Wildenberg, Kasthuri Lab, UChicago

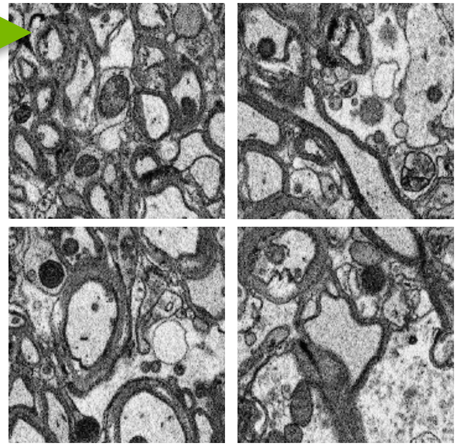
SAMPLE PROCESSING

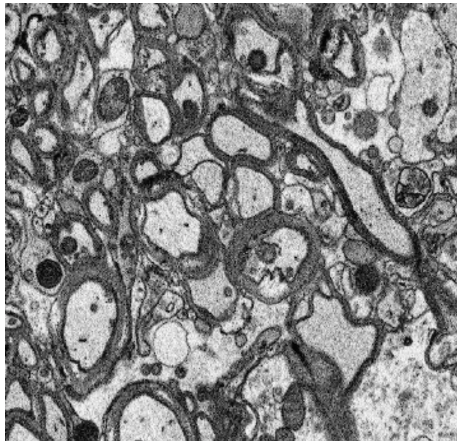
Mouse brain: 70M neurons



sample

section
section
section
section
section
section





~1cm^3

~1mm^3

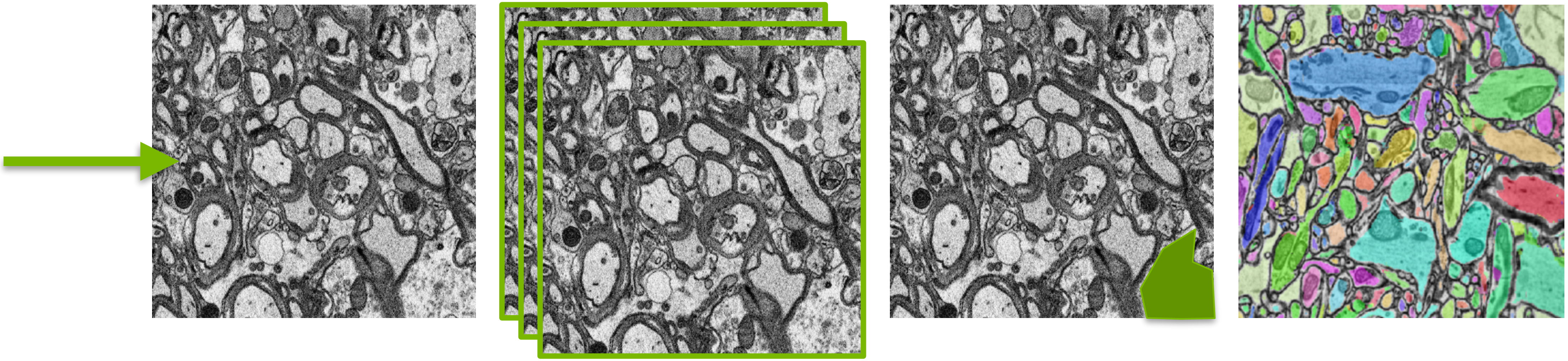
25000
40nm sections
1mm x 1mm
(6nm resolution)

Each section imaged with EM as N tiles (8 bit)

Sections stitched together

How much image data is 1mm^3 ? 1e15 voxels -> ~1 PB

SAMPLE PROCESSING

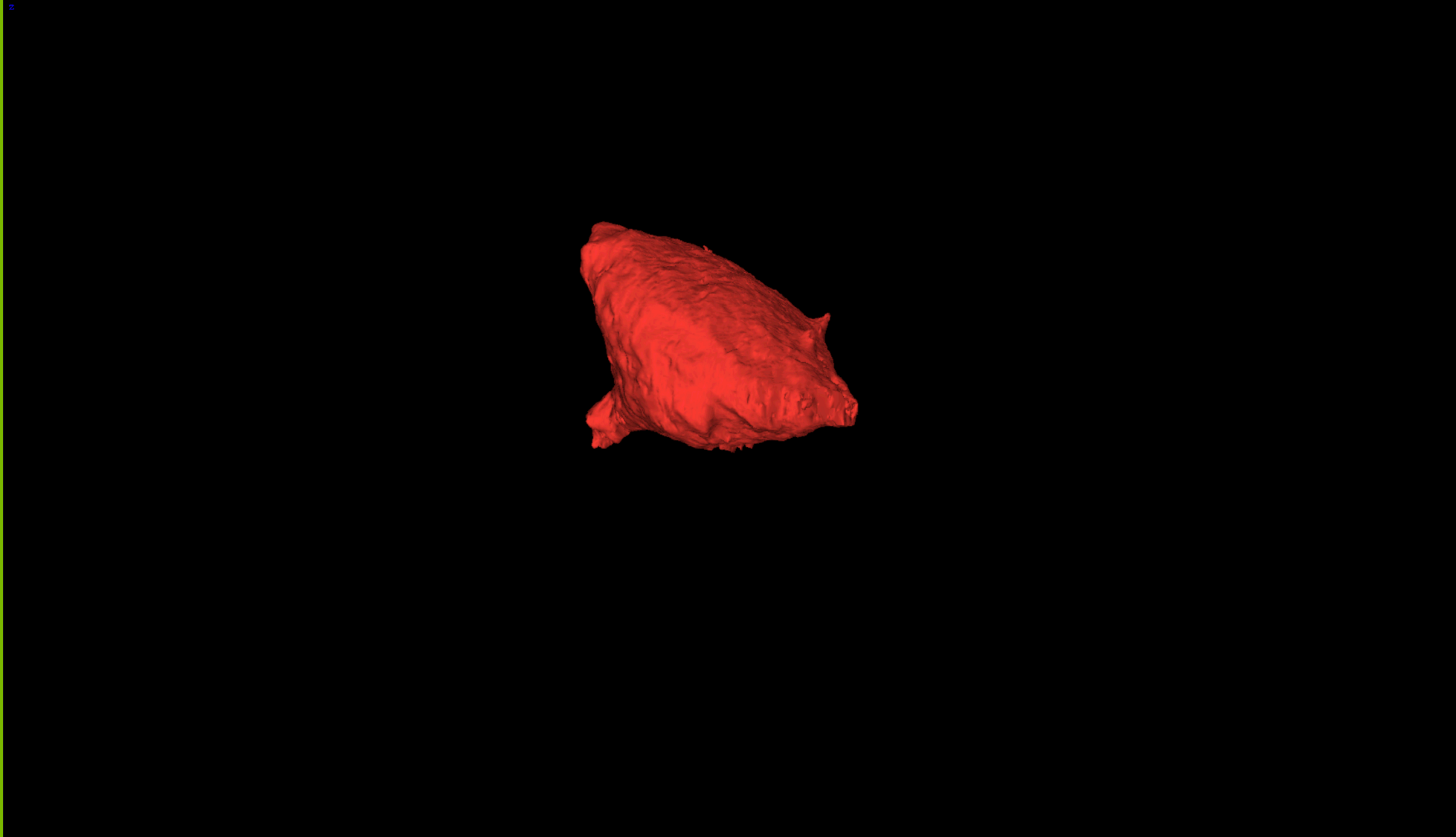


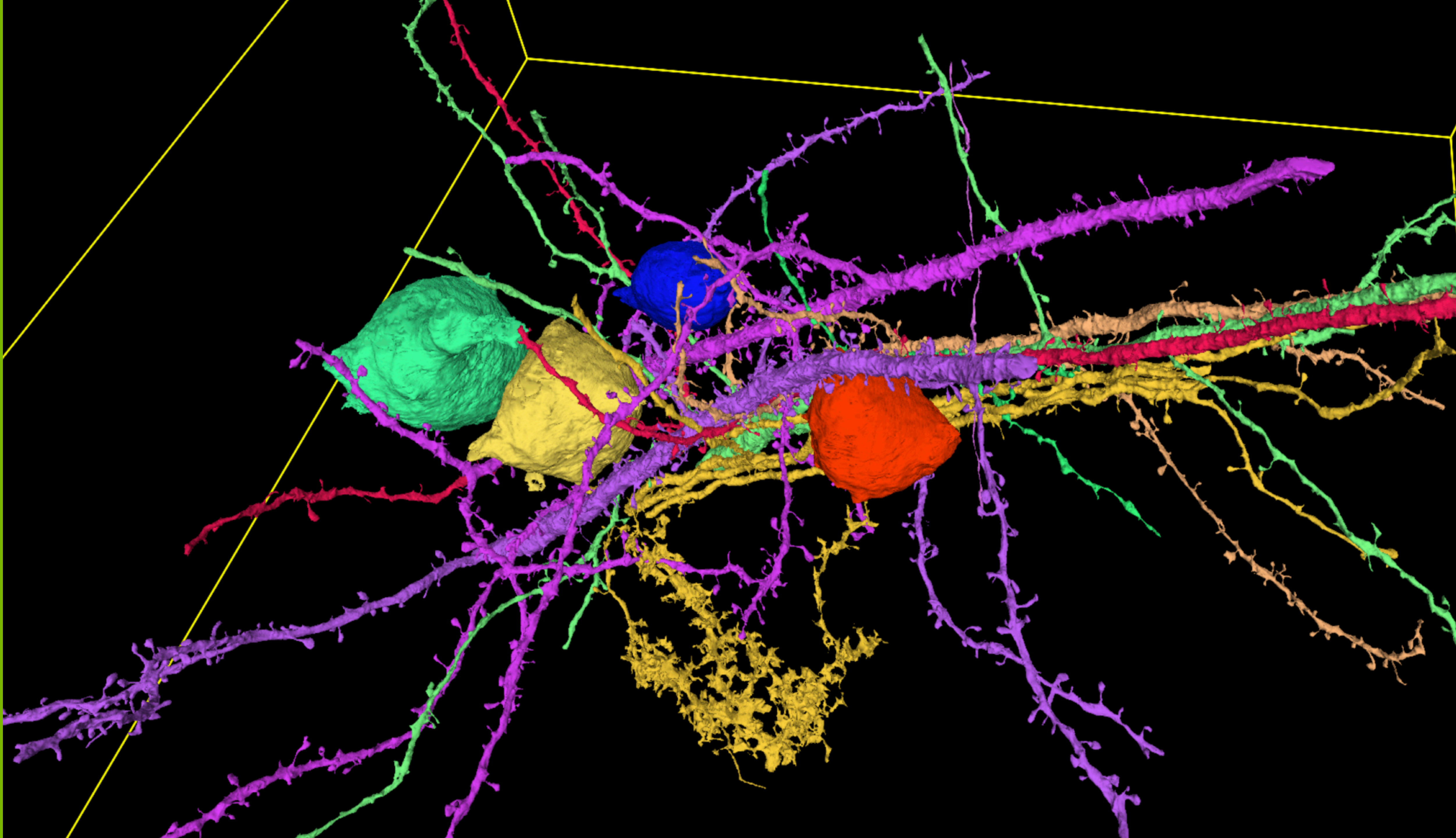
Sections stitched together

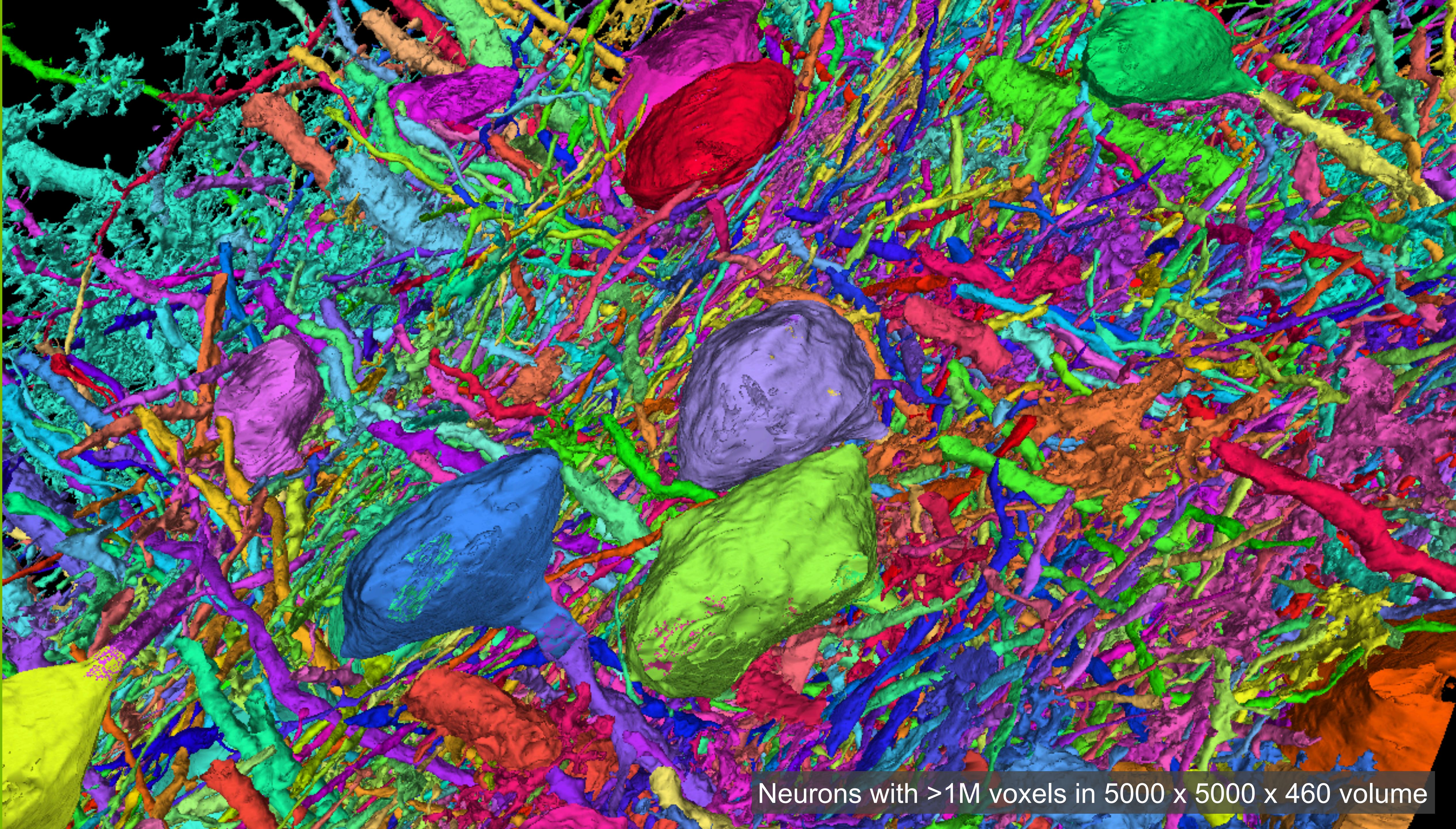
Align sections

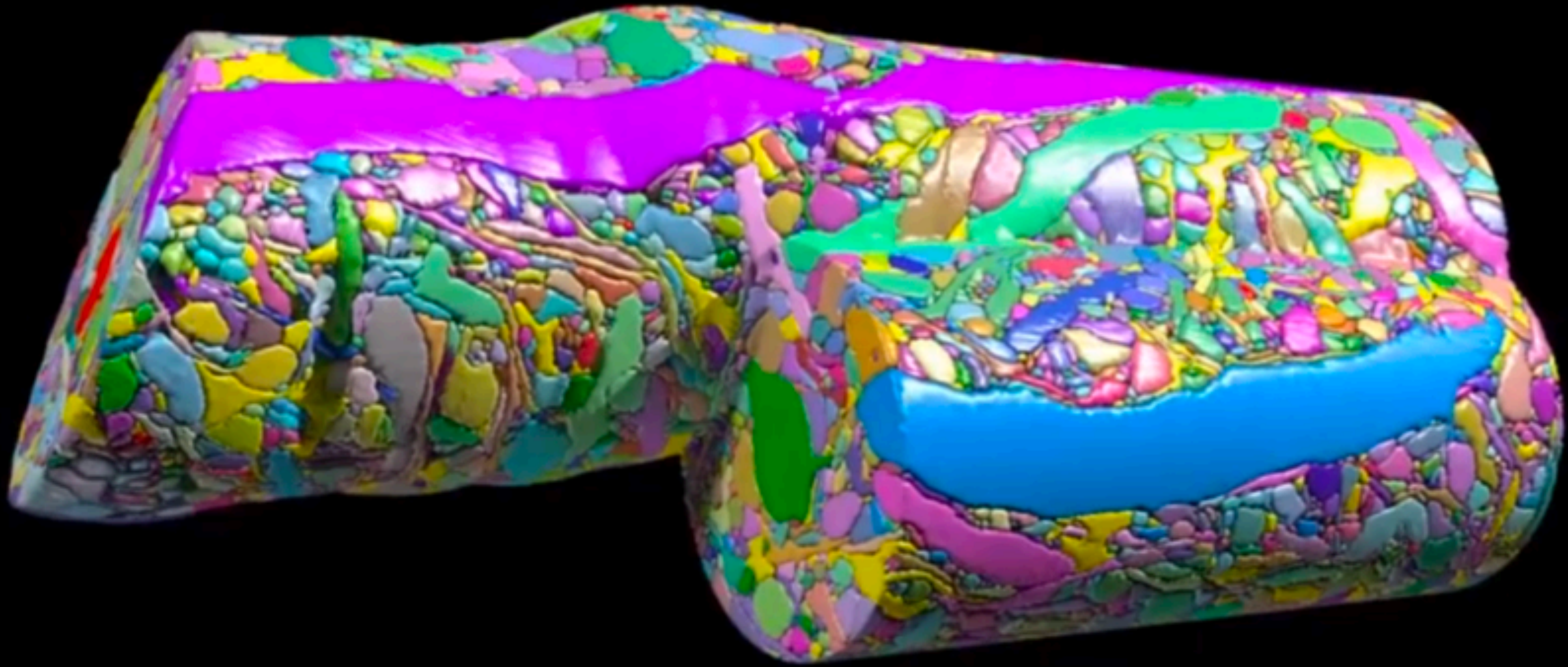
Mask out non-target objects

Segment target objects


z

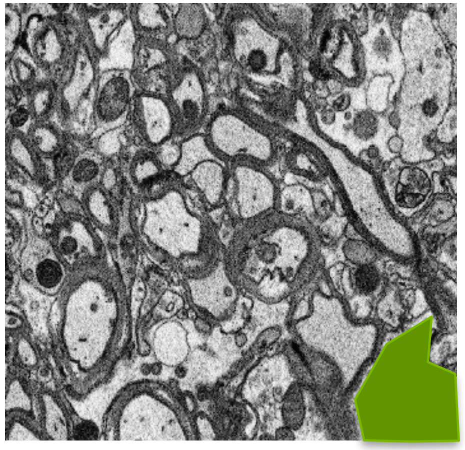



Neurons with >1M voxels in 5000 x 5000 x 460 volume



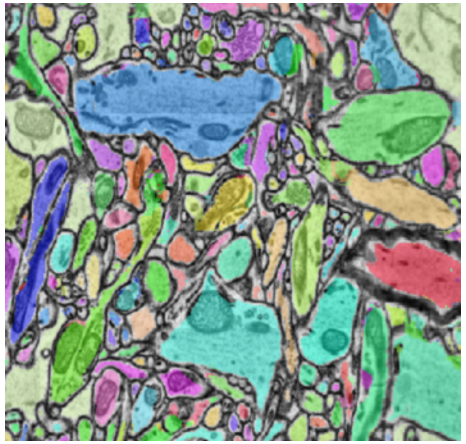
Kasthuri N, Hayworth KJ, Berger DR, Schalek RL, Conchello JA, Knowles-Barley S, Lee D, Vázquez-Reina A, Kaynig V, Jones TR, Roberts M, Morgan JL, Tapia JC, Seung HS, Roncal WG, Vogelstein JT, Burns R, Sussman DL, Priebe CE, Pfister H, Lichtman JW. Saturated Reconstruction of a Volume of Neocortex. Cell. 2015 Jul 30;162(3):648-61. doi: 10.1016/j.cell.2015.06.054. PMID: 26232230.

LEARNING CHALLENGES IN CONNECTOMICS



Mask out non-target objects (e.g. blood vessels)

2D Unet used to segment large objects and mask them from later expensive segmentation operations

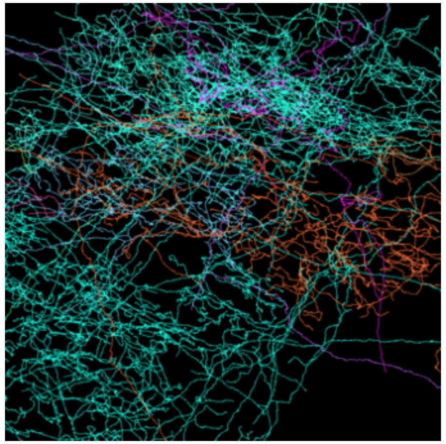


Segment target objects

Flood Filling Network (FFN) used to segment neurons incl their axons, dendrites, dendritic spines. **Training is human-intensive; inference is compute-intensive.**

Human-intensive annotation to produce training data and iteratively correct segmentation

SOTA efforts require **thousands of hours** of human effort to annotate and correct a small fraction of a mouse brain



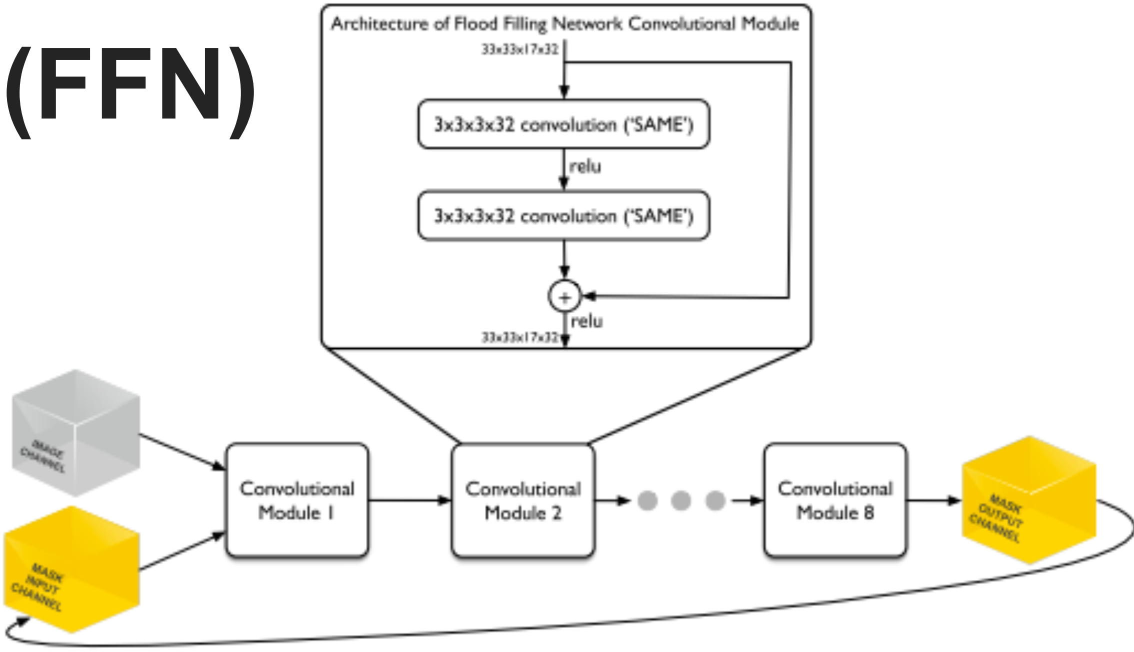
Connectivity Graph Analysis

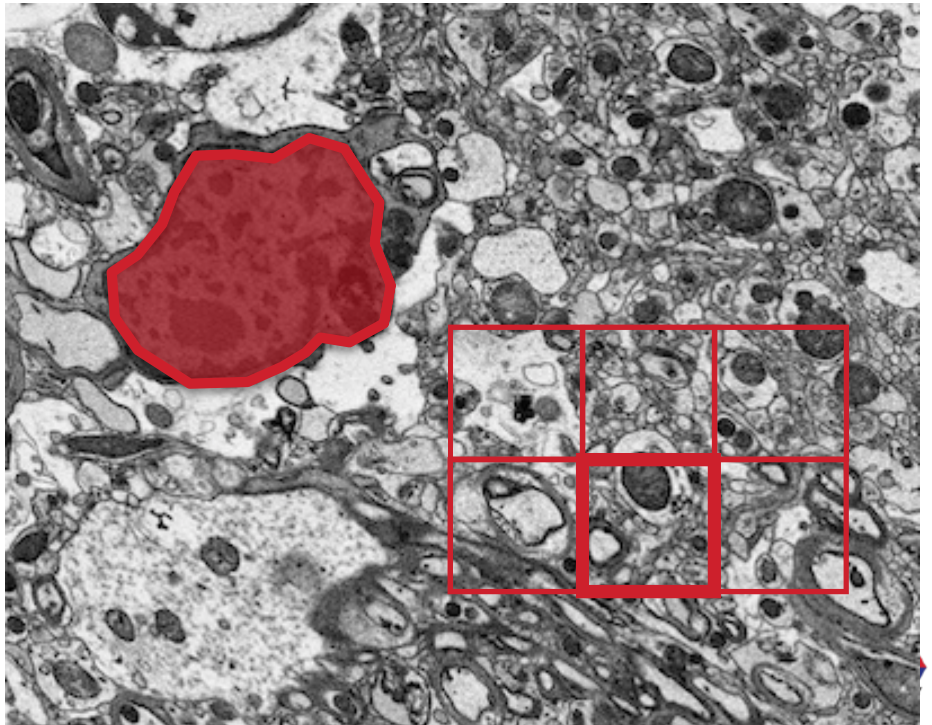
3D Unet used to segment synapses (connections between neurons) in the final 3D segmentation yields opportunity to analyze interconnections

FLOOD FILLING NETWORK (FFN)

TRAINING

- FFN was developed by Google Brain for automated segmentation of structures in EM image data
- Implemented in TensorFlow
- Network consists of a series of 3D convolutional blocks with residual connections
- A 12 layer network has roughly 0.5M trainable parameters
- Network builds on notion of watershed algorithm
  - Finds boundaries and fills interior
  - Prefers split errors over merge errors
- Boundary-finding is complicated
  - complex object structure and substructure
  - variation between datasets (limited opportunity for transfer learning, and yet...)
  - variation within datasets (fixing, cutting, staining, imaging)
- Dataset-specific training is required
- Accuracy of FFN is an order of magnitude better than past approaches, but with a higher computational cost.



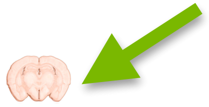


M. Januszewski, J. Kornfeld, P. H. Li, A. Pope, T. Blakely, L. Lindsey, J. Maitin-Shepard, M. Tyka, W. Denk, and V. Jain, "High-precision automated reconstruction of neurons with flood-filling networks," Nature Methods

DATA CHALLENGES IN CONNECTOMICS

Mouse brain: 70M neurons

tiny brain



~1cm^3

How much image data is 1cm^3 ? **~1EB**

How many brains do we need for statsig?

Human brain: 80B neurons

~1000cm^3

How much image data is 1000cm^3 ? **~1000 EB**

(6nm x 6nm x 40nm)

Reconstructed data will be much larger:

- Segmentation labels for each voxel (4x voxel data)
- 3D Mesh
- Skeleton

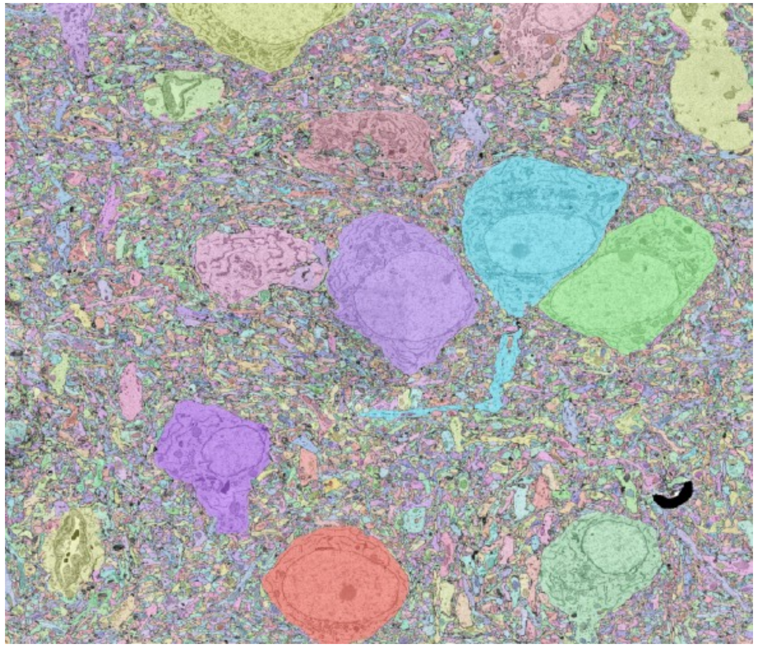
Need tools for navigating these datasets

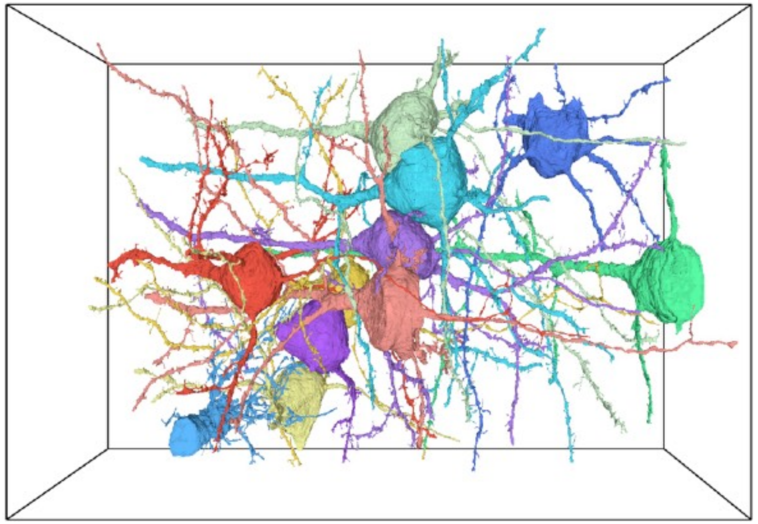
Argonne NATIONAL LABORATORY

LARGE-SCALE RECONSTRUCTION

Theta: 262K cores, million-way concurrency

- **Image stitching** on many nodes with existing codes; scalable code in development (parallel granularity: sections or tiles)
- Parallelizing existing **alignment** codes to run at large scale (SEM and MSEM); scalable code in development (parallel granularity: section pairs or region pairs)
- **Scaling FFN training and inference** to thousands of nodes on Theta (TensorFlow+Horovod) (parallel granularity: overlapping subvolumes)
- Ran initial hyperparameter scans on FFN to explore performance (batch size, learning rate, optimizer, number of convolutional layers)
- **Hyperparameter optimization** with Deep Hyper (ongoing)
  - deephyper.readthedocs.io
- Exploring variations of FFN network architecture
- Application of learning in other parts of the pipeline (e.g. image alignment)
- **For Aurora, working with Intel** to optimize codes for future compute architectures
- Evaluation of impact of reduced precision on accuracy and runtime
- Deployed web-based tools for remote annotation (webKnossos) and visualization (neuroglancer) backed by ALCF storage

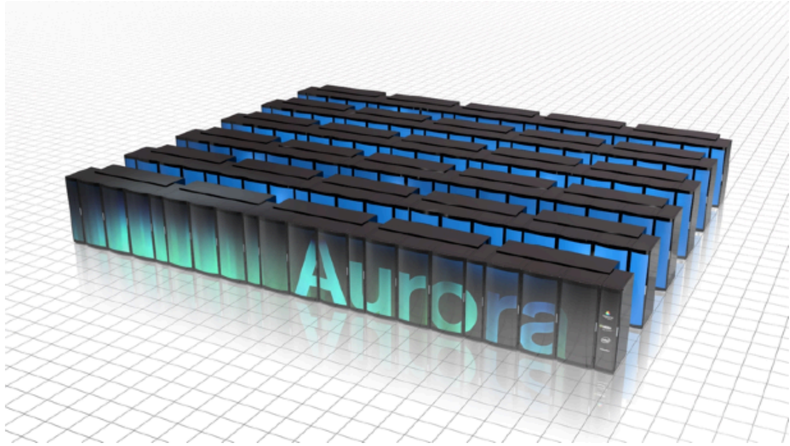




Dong, et al, "Scaling Distributed Training of Flood-Filling Networks on HPC Infrastructure for Brain Mapping", 2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS) at SC19

Vescovi, et al, "Toward an Automated HPC Pipeline for Processing Large Scale Electron Microscopy Data", 2020 IEEE/ACM 2nd Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP) at SC19

CONNECTOMICS+HPC FUTURE



Significantly larger scales lie ahead

- Connectomics data: Whole mouse brain; many mouse brains; human brain
- Imaging tech: Many-beam electron microscopy; X-ray imaging (APS)
- Supercomputing: Exascale computing in 1-2 years

Rough projection from today's reconstruction to future whole mouse brain

- 6nm x 6nm x 40nm resolution: ~1 exavoxels
- Sample prep alone will require significant time/many EMs
- Segmentation time on today's system **Theta: 50+ years**
- Segmentation time on future system **Aurora: ~3 years**
  - Worth noting: connectomics will unfortunately not get exclusive access to Argonne's exascale supercomputer
  - **Essential that we improve performance of current methods and develop new methods to accelerate all stages of the reconstruction pipeline**